The Quality of Stakeholder-Based Decisions: Lessons from the Case Study Record

Thomas C. Beierle

November 2000 • Discussion Paper 00–56

Resources for the Future
1616 P Street, NW
Washington, D.C. 20036
Telephone: 202–328–5000
Fax: 202–939–3460
Internet: http://www.rff.org

The Quality of Stakeholder-Based Decisions: Lessons from the Case Study Record

Thomas C. Beierle

Abstract

The increased use of stakeholder processes in environmental decisionmaking has raised concerns that the inherently "political" nature of such processes may sacrifice substantive quality for political expediency. In particular, there is concern that good science will not be used adequately in stakeholder processes nor be reflected in their decision outcomes. This paper looks to the case study record to examine the quality of the outcomes of stakeholder efforts and the scientific and technical resources stakeholders use.

The data for the analysis come from a "case survey," in which researchers coded information on over 100 attributes of 239 published case studies of stakeholder involvement in environmental decisionmaking. These cases reflect a diversity of planning, management, and implementation activities carried out by environmental and natural resource agencies at many levels of government.

Overall, the case study record suggests that there should be little concern that stakeholder processes are resulting in low quality decisions. The majority of cases contained evidence of stakeholders improving decisions over the status quo; adding new information, ideas, and analysis; and having adequate access to technical and scientific resources. Processes that stressed consensus scored higher on substantive quality measures than those that did not. Indeed, the data suggested interesting relationships between the more "political" aspects of stakeholder decisionmaking, such as consensus building, and the quality of decisions.

Key Words: public participation, stakeholder, science, alternative dispute resolution, consensus building

Contents

The Quality of Stakeholder-Based Decisions: Lessons from the Case Study Record

Thomas C. Beierle*

1. Introduction

Stakeholder participation in environmental decisionmaking has increased at all levels of government in the last decade. Among federal agencies, the Environmental Protection Agency (EPA), Department of Energy (DOE), and Department of Defense have initiated over 200 citizen advisory groups at contaminated sites around the country (FFERDC, 1996). State environmental agencies conducting comparative risk projects have convened interest group representatives and the general public to help make decisions about environmental priorities (Perhac, 1998; WCED, 1997). Local governments have increasingly engaged citizens in watershed management activities, sustainability projects, and a myriad of other planning and management activities.

Underlying much of the move toward greater stakeholder involvement is a recognition that environmental decisions are "political" as well as scientific. That is, resolving environmental problems requires addressing the interests and values of the public in ways that cannot be resolved with science alone. A focus on "consensus building" and "alternative dispute resolution" in many stakeholder processes is an explicit effort to accommodate the political aspects of environmental decisionmaking.

Some analysts have raised the question, however, of whether stakeholder processes are shifting the emphasis of environmental decisionmaking too far in the political direction. They are concerned that stakeholder processes may sacrifice the quality of decisions—and scientific and technical quality in particular—in pursuit of political expediency. In a recent examination of stakeholder processes, Yosie and Herbst (1998) stated that scientists and scientific information are typically not well-integrated into stakeholder decisionmaking. Gregory (2000) recently

* The author would like to thank Jerry Cayford, David Konisky, and Anne McEnany for their assistance in creating the public participation case survey data set. Jerry Cayford also provided very insightful comments on various drafts of this paper. The author would also like to thank Terry Davies, Allen Blackman, Mike Toman, Allan Mazur, Gail Charnley, and Caron Chess for their reviews and helpful comments. The research on which this paper is based was supported by the National Science Foundation's Program in Decision, Risk, and Management Science under award number 9818728.

suggested the need to de-emphasize consensus in stakeholder processes in favor of competent problem-solving aimed at producing high quality, but perhaps not universally accepted, decisions. EPA's Science Advisory Board (SAB) recently expressed its concern about the quality of stakeholder decisions in a letter to EPA Administrator Carol Browner (Daisey and Morgan, 1999). In the letter, the SAB queried whether stakeholder processes at EPA were moving the agency away from the use of good science, perhaps leading it to abrogate its responsibility to pursue the public interest.

Concerns about the role of science in stakeholder-based decisions are a subset of a larger set of issues about the substantive quality of decisions made by stakeholders. Although there is much in the public participation literature about the quality of stakeholder processes, there is very little about the quality of decisions that stakeholders actually make. In fact, in their review of stakeholder processes, Yosie and Herbst (1998, 49) noted that there was broad disagreement in the environmental policy community over whether stakeholder involvement improved decisions or not.

To settle the disagreement over the quality of stakeholder-based decisions, it would be ideal to compare—using a common metric—a set of decisions made through stakeholder involvement with a similar set of decisions made without stakeholder input. However, finding an opportunity to conduct such a natural experiment on more than a handful of case studies is unlikely. To date, comparative studies have only been successful for the more exotic manifestations of stakeholder participation. Coglianese (1997), for example, used a comparative approach to evaluate regulatory negotiations, but the outcomes of interest were speed and conflict resolution rather than decision quality. Even if a broader opportunity for comparison presented itself, the question of whether stakeholders contribute to or detract from decision quality would unlikely be settled, because there is no agreement on what "quality" means. Should it account for the multiple objectives of stakeholder participation, such as capacity building and conflict resolution? Should it account for diverse sets of public values?

In this paper, we seek to shed light on the quality of stakeholder-based decisions by departing from the comparative ideal. We evaluate 239 case studies of public participation in environmental decisionmaking against a broad set of rather conventional quality criteria. Data for the analysis were derived from published case studies on 239 cases of public participation. Researchers coded over 100 attributes of each case related to its context, the participatory process used, and the outcomes. To a great extent, the cases examined here do not concern high-profile, federal-level decisionmaking. Most deal with more routine cases of planning,

management, and implementation carried out by environmental and natural resource agencies at many levels of government.

The criteria used to evaluate the cases were as follows:

1. Were decisions more *cost-effective* than alternatives?
2. Did decisions increase *joint gains* over alternatives?
3. Was it the *opinion* of participants or authors that decisions were improved?
4. Did *other measures* suggest improvements in quality?
5. Did stakeholders *add information*?
6. Did stakeholders contribute to the *technical analysis* of problems?
7. Did stakeholders generate *innovative ideas*?
8. Did stakeholders introduce a more *holistic perspective*?
9. Did stakeholders bring *technical capacity* to the process?
10. Was there adequate *access to information and expertise*?

The data derived from these criteria were aggregated into two main quality measures. The first measure dealt with the outcomes of stakeholder processes and pulled together the first eight criteria. It reflected a broad definition of quality that incorporated both "technical" considerations, such as cost-effectiveness, and more "political" considerations, such as increasing joint gains. The second measure dealt with process and brought together the last two criteria. It reflected a more specific definition of quality—one focused on scientific and technical quality. Given the difficulty of identifying an outcome measure that can somehow measure "scientific quality," these process criteria may provide the best perspective we have on the specific issue of the scientific and technical quality of stakeholder-based decisions.

The results of the analysis indicate that there should be little concern that stakeholder processes, as viewed through the case study record, are resulting in low quality decisions. Across a broad range of criteria, the majority of cases contained evidence of decisions that were better than alternatives or evidence that stakeholders added new information, ideas, and analysis. In most of the cases, numerous technical and scientific resources were available to stakeholders, either through their own training or through outside expertise. The analysis also showed that stakeholder processes that pursued consensus scored higher on quality criteria than processes that

did not seek consensus. Finally, the nature of the lead agency seemed to bear little relation to the quality of stakeholder-based decisions.

Section 2 of this paper discusses the methodology used to collect and analyze the data on stakeholder processes. Section 3 outlines the pool of 239 case studies, describing the range of environmental issues covered, the level of government of lead agencies, and the types and characteristics of participatory processes used. Section 4 turns to the evidence regarding the substantive quality of stakeholder-based decisions. Section 5 examines questions about how the substantive quality of decisions is related to the participatory process and the nature of the lead agency. Section 6 concludes the paper with a brief wrap-up of results and a few issues for further consideration in thinking about the quality of stakeholder-based decisions.

2. Methodology and Data

The data for this paper come from a "case survey" of 239 cases of public participation in environmental decisionmaking. A case survey is analogous to a normal closed-ended survey, except that a "reader-analyst" "asks" a standard set of questions of a written case study rather than of a person (Lucas, 1974; Yin and Heald, 1975; Bullock and Tubbs, 1987; Larrson, 1993). It is a formal process for systematically coding relevant data from a large number of qualitative sources for quantitative analysis. Derived data can support data analysis even if the questions addressed in the analysis are different from those posed in the original case study (GAO, 1991).

Researchers screened over 1,800 case studies—drawn from journals, books, dissertations, conference proceedings, and government reports—ultimately identifying the 239 cases making up the data set.[1] Each case was coded for over 100 attributes covering the type of environmental issue, the people who participated, important features of the participatory process, and the outcomes achieved. Each attribute was assigned a score—usually low, medium, or high—based

[1] Case studies were screened based on the following criteria:

- dealt with public involvement in environmental decisionmaking, generally at the administrative level;
- occurred in the United States;
- occurred since 1970;
- had an identifiable lead (or otherwise interested) government agency;
- described a discrete mechanism (or set of mechanisms) used to engage the public;
- described participation of nongovernmental citizens other than regulated parties; and
- contained sufficient information on context, process, and outcomes.

on a standard template. Each score was given one of three weight-of-evidence measures, ranging from "solid evidence" to "best informed guess." Data with the lowest weight of evidence were not used in the analysis. The scores were accompanied by a written justification that recorded important qualitative information.

Each case was coded by one of three researchers or by pairs of them. In order to ensure consistent coding among researchers, a process of inter-coder reliability testing was used. This involved pairs of researchers reading and coding the same subset of case studies independently and then comparing codes. Where there were conflicting codes, they were resolved through consensus and the coding template was clarified as necessary. Pairs of researchers continued to code sets of cases in parallel until they consistently achieved two-thirds agreement, a level of reliability regarded as satisfactory in the literature (Larsson, 1993). Each case was then coded by only one researcher. As the coding progressed, researchers would periodically code a set of cases in parallel to assure that inter-coder reliability was being maintained. Overall, around 10% of the cases were used in the inter-coder reliability process.

Data analysis consisted mainly of counts of scores and a review of the qualitative information accompanying them. Relatively simple comparative statistics were used to develop correlation coefficients and to identify statistically significant differences between sets of data. The statistical analysis used a Kendall's tau-b correlation coefficient and a chi-squared test of significance.[2]

Although it has been used in the policy analysis and business literature, the case survey methodology is still somewhat experimental, and there are a few important caveats to mention. The quality of the data used in a case survey is only as good as the quality of the case studies from which the data come. Moreover, cases by different authors and for different purposes will report on different aspects of a process, leaving some data gaps. The analysis accounted for these problems somewhat by the assignment of weight of evidence scores and by drawing on enough cases to overcome problems with data gaps.

---

[2] The Kendall's tau b correlation coefficient is based on the number of concordant and discordant pairs of observations in a contingency table, using a correction for ties. Its calculation is described in STATA (1997, p. 487). It is an appropriate non-parametric measure of correlation for ordinal data (Bullock and Tubbs, 1987, p. 210). A rule of thumb for using the chi-squared test of association is that the expected count of each cell in the contingency table should be greater than 5 (and preferably greater than 10), which was met in most cases here (Stokes et al., 1995).

Perhaps more important is whether the data set is biased. Of particular concern is a success bias—that only successful cases were written up and that authors had a tendency to overemphasize the good things that stakeholders accomplished. We analyze these potential sources of bias explicitly in Appendix B. More generally, there are two reasons to think that a success bias may not be as prevalent in the published literature on stakeholder participation as it is in other research literatures, such as the hard sciences. First, authors don't necessarily have an incentive to write up only successful cases. Many of our cases came from doctoral dissertations or other studies where multiple cases were compared and unsuccessful cases provided as much, or more, insight as successful ones. Second, different authors defined the "success" of stakeholder processes differently, and very few defined it in terms of decision quality. Even if there is an overall success bias, then, it is unlikely to extend to the quality criteria used here. Overall, bias does not appear to have much impact on the main conclusions of this paper.

3. Overview of the Case Studies

In discussing results, it is important to know just what kind of participation is being talked about. The bulk of public-participation cases covered here are not those that make newspaper headlines. They concern a diversity of planning, management, and implementation activities carried out by environmental and natural resource agencies at many levels of government.

More than 80% of the cases dealt with decisions that were specific to a single site or geographic feature, such as cleaning up contaminated property, siting an industrial facility, or developing a management plan for a park. The remaining 20% of cases dealt with broader policy issues, such as the development of regulations or the identification of environmental protection priorities for a state or region. Forty-four percent of the cases dealt with pollution-related issues and the rest concerned natural resources, such as wildlife, forests, mining, and agriculture. Figure 1 identifies the distribution of cases among broad categories of issues.

In 55% of the cases, state and local agencies took the lead, with cases covering 40 states (see Figure 2). Most of the remaining cases were led by federal agencies, primarily EPA, DOE, the Army Corps of Engineers, and various resource management agencies within the Department of the Interior. In nearly three-quarters of the cases, the lead agency was actively overseeing or participating in the process. In the remaining cases, the lead agency had either delegated oversight to another organization or was simply part of the intended audience of efforts initiated outside of government altogether.

The types of processes in which stakeholders participated ranged from public meetings to intensive negotiations (see Figure 3). These processes defy easy categorization. However, some of their major differences can be captured with reference to a few design features: the nature of the typical participant, the method used to reach decisions, and the nature of the intended outcomes, as described below.

Twenty-one percent of the cases involved public hearings, meetings, and workshops. In most of these cases, access was open: any interested citizen could participate. While participants might identify with major interest-orientations—pro-environment, pro-business, anti-tax, and so forth—or be members of interest groups, their role was that of individual citizens, not of formal representatives of some group. These processes mainly involved information exchange, with agencies informing citizens about their activities and citizens providing input and individual opinions on agency policy. Agencies were under an implicit obligation to review information from these processes, but in most cases there was little commitment to actually share decisionmaking authority with the public.

Fifty-six percent of the cases concerned the work of advisory committees. Unlike public meetings, advisory committees typically had a defined and consistent membership. In most cases, participants were selected to represent various interest groups or points of view, although in a few cases they were selected to be "representative"—that is, a microcosm of the socioeconomic characteristics and issue-orientation of a particular area. In contrast to public meetings, these processes were as much about the interaction among participants—who frequently brought very different views on a decision to the table—as about providing input to a lead agency. The work of advisory committees typically took place in ongoing, regular meetings, some of which stretched out over years. Typically the outcome of advisory committee work was a set of recommendations to a lead agency.

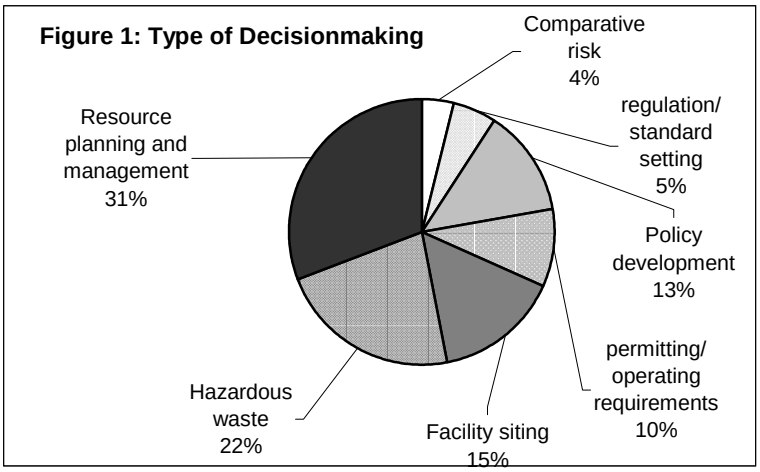


Figure 1: Type of Decisionmaking
Comparative risk 4%
regulation/ standard setting 5%
Policy development 13%
permitting/ operating requirements 10%
Facility siting 15%
Hazardous waste 22%
Resource planning and management 31%

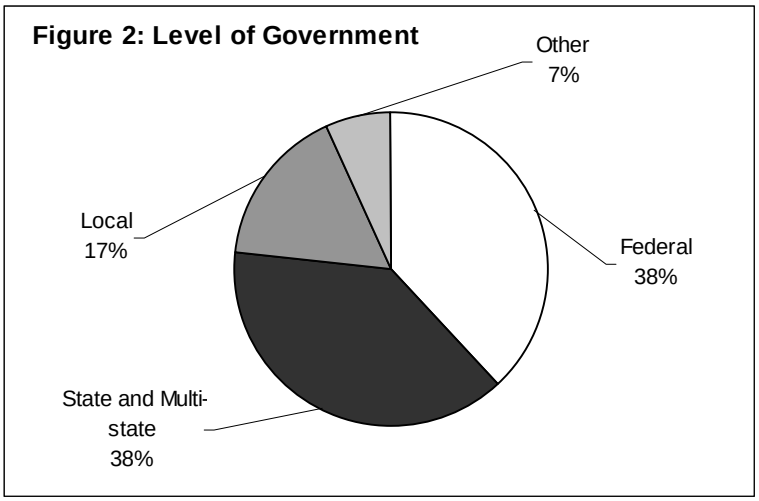


Figure 2: Level of Government
Other 7%
Federal 38%
State and Multi-state 38%
Local 17%

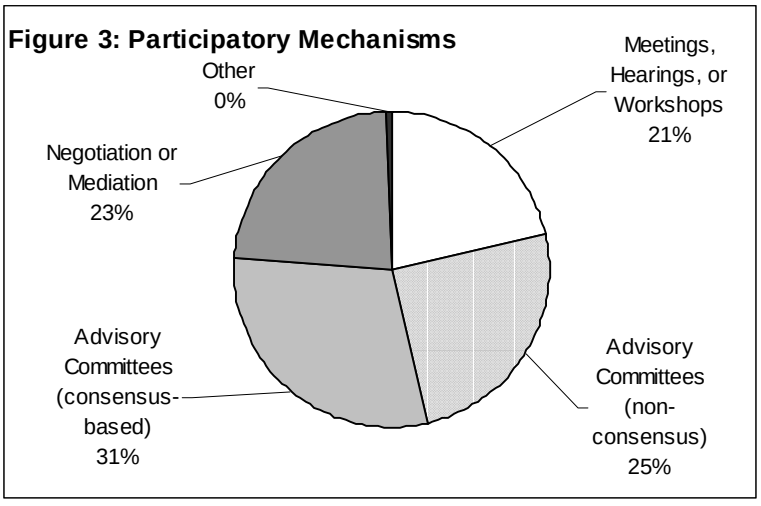


Figure 3: Participatory Mechanisms
Meetings, Hearings, or Workshops 21%
Advisory Committees (non-consensus) 25%
Advisory Committees (consensus-based) 31%
Negotiation or Mediation 23%
Other 0%

A major distinction can be made among advisory committees based on whether they operated by consensus or not. In the cases using consensus, decisionmaking took on aspects of internal negotiations among participants, often complete with facilitation by a third party. In about half of the advisory committee cases, consensus was eschewed in favor of voting or the presentation of competing sets of recommendations. The other half of the advisory committee cases used consensus, forcing opposing interests to work together to come to a common and acceptable solution to a problem.

The final 23% of cases dealt explicitly with negotiations and mediations. In negotiations and mediations, unlike the public meetings or advisory committees, stakeholders were actually formulating agreements that would bind their organizations to particular courses of action. In some cases, the negotiating parties themselves implemented the agreement, as was the case with many watershed management groups, for example. In other cases, parties agreed to be bound by a decision in exchange for a strong commitment that a lead agency would act on it. The participants in a negotiation or mediation were typically professional representatives of organized interest groups or other entities. They spoke for the views of those they represented and made commitments on their behalf. By definition, decisions were made by consensus.

Regardless of the participatory process, the level of government of the lead agency, or the topics under discussion, in all of these cases stakeholders had some direct or indirect role in affecting the quality of environmental decisions ultimately made. The next section discusses the criteria used to examine the quality of stakeholder-based decisions and what the cases show about it.

4. The Quality of Stakeholder-Based Decisions

Posing questions about the quality of stakeholder-based decisions raises important and difficult issues about the purpose and appropriate evaluation of stakeholder processes. When confronted with the myriad motivations for bringing stakeholders to the table, some of the traditional criteria for assessing quality—such as cost-effectiveness or improved information—appear to be quite narrow measurement tools. Stakeholder processes have many and varied purposes beyond making decisions, such as capacity building and social learning, conflict resolution, and networking. Perhaps more importantly, the adoption of a stakeholder process is an implicit acknowledgement that the environmental decision in question has important value-dimensions that are not captured by traditional quality measures. In fact, many analysts of public

participation have eschewed evaluation of decision quality, because defining "quality" is so value-laden.

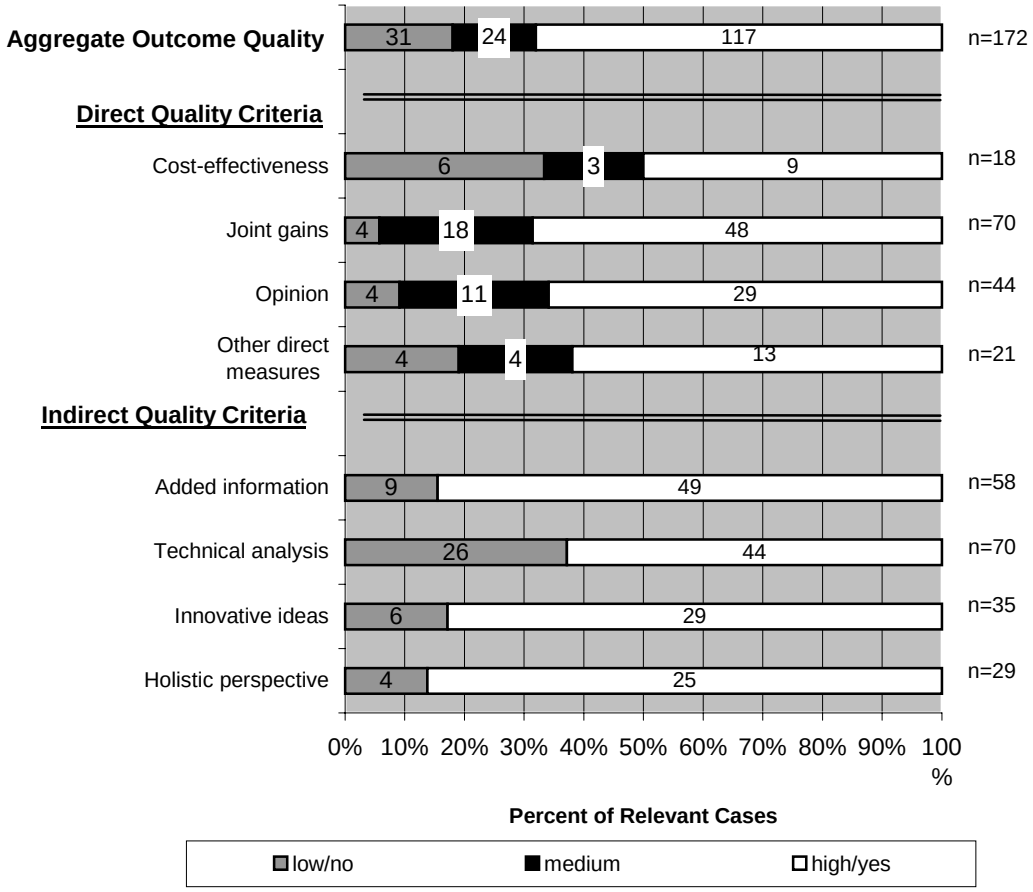
Although the reluctance in the literature to address the quality dimension of stakeholder decisions is understandable, it has left a void in knowledge about how the move toward stakeholder decisionmaking is affecting environmental policy. There are, in fact, a number of measures of quality—however imperfect or incomplete—that can be applied. We can distinguish two sets of quality criteria. The first examines the substantive quality of outcomes. The second examines the process, focusing on the technical and scientific resources available to stakeholders. Each is discussed below along with the results of the data analysis.

Substantive Quality of Outcomes

To measure the substantive quality of outcomes we use an aggregate of eight separate, but related, quality criteria. It is appropriate to use a variety of criteria here because different kinds of processes will affect decision quality in different ways. An agreement developed through mediation, for example, can be evaluated against a likely alternative. The contributions that citizens make at a public meeting, however, require criteria that look to these contributions specifically, rather than to the decisions ultimately made.

Figure 4 shows the results, across cases, for the eight individual criteria and for an aggregate outcome quality measure, which runs along the top of the figure. (Details on how the aggregate measure was constructed are contained in Appendix A.) As shown in the figure, 172 cases out of a total of 239 were scored for the aggregate outcome quality measure. There were roughly twice as many high scoring cases as low- and medium-scoring cases combined. This means that in roughly two-thirds of the cases for which information was available, evidence—sometimes from multiple component criteria—suggested that outcomes were of high substantive quality. To better understand what substantive quality means in these cases, we turn below to a more complete explanation of the eight component quality criteria, four of which evaluate quality directly and four indirectly.

Figure 4: Outcome Quality



Notes:
1) Numbers within bars refer to the number of cases in that category
2) Direct quality criteria were measured as high/medium/low; indirect quality criteria were measured as yes/no and translated into high/low for the aggregate.

The direct measures rank decisions along a single quality dimension: decision A is better than decision B according to criterion X. Some of the more concrete direct measures come out of the program evaluation literature, where they are used as proxies for the extent to which the public interest (however interpreted) is being served. Others are less specific. In each case, the direct quality criteria were scored as high, medium, or low based on how well the criterion was met in a particular case.

The direct criteria are most applicable to cases where decisions were actually made by stakeholders, either in the form of group recommendations to an agency or actual agreements among participants on a course of action. Most of these cases were scored based on comparisons that participants or case study authors implicitly made between stakeholder decisions and likely alternatives. Such comparisons provided a case-specific baseline—albeit often a rather vaguely defined one—against which the stakeholder decisions could be scored.

The direct quality measures used were:

Cost-effectiveness. Were the decisions or recommendations made by participants more or less cost-effective than a probable alternative? This criterion does not refer to the cost-effectiveness of undertaking a stakeholder process relative to some other approach to making decisions, but looks at the decision itself. For example, the DOE credited the Fernald Citizens Task Force—a stakeholder advisory committee established to advise DOE on the remediation of its Fernald, Ohio, nuclear weapons facility—with designing a cleanup plan that saved taxpayers over $2 billion over the life of the project (Applegate, 1998). Only 18 cases could be scored for cost-effectiveness. For these, the number of high-scoring cases matched the number of medium- and low-scoring cases combined.

Joint Gains. Were some participants made better off through agreement without any participant becoming worse off? This is a standard measure in the negotiation literature that traces its roots to measures of "Pareto optimality" in game theory and the economics of Coasian bargaining. In an analysis of environmental mediation, Bingham (1986) used agreement among parties as a proxy for joint gains, arguing that if they could get a better deal somewhere else they would defect from the mediation. The joint gains criterion was coded for more cases than any other direct quality criterion, with 48 cases showing an improvement, 18 showing no change, and only 4 showing a decrease in quality.

Opinion. Did participants or case study authors feel that decisions were better than a reasonable alternative? Often quality was not expressed according to concrete criteria but as general

satisfaction with an outcome or in terms of a range of quality criteria. For example, in the Woodburning Stove Emissions regulatory negotiation, one participant said that the regulations developed by the group were "more effective, more environmentally-oriented, fairer to the industries, and more creative than those EPA could have been expected to develop." Of the 44 cases scored for this criterion, 29 were high, 11 medium, and 4 low.

Other measures. Were decisions deemed substantively better due to some other measure of quality? All other direct measures of quality mentioned by case study authors were collected here. They encompassed a variety of criteria, such as the scientific foundation of the decision, its technical quality, or whether it resulted in a more environmentally beneficial outcome. For example, some participants in the development of a habitat management plan for Clark County, Nevada, "acknowledged that at times the biological ideal was compromised, but the best possible outcome was achieved." The compromise in the biological ideal "at times" earned this case a medium score for this criterion. Of the 21 cases scored for this criterion, 13 were high, 4 were medium, and 4 were low.

The four indirect quality measures look less at the overall quality of decisions than at what stakeholders brought to decisions. Did stakeholders add new information, ideas, or resources to the decisionmaking table that would not otherwise have been available? The indirect criteria were more appropriate for examining cases where the public was not actually making decisions, but was contributing to the knowledge base that government agencies would then use to determine a course of action. In scoring these cases, we assumed that some public input of information, ideas, or analysis was better than no public input at all. The indirect quality criteria were scored as "yes" or "no" based on whether the criterion of interest was met in the case or not. To construct the aggregate outcome quality score, "yes" scores were interpreted as "high," and "no" scores were interpreted as "low" as detailed in Appendix A. The indirect criteria are as follows:

Added information. Did participants add information to the analysis that would not otherwise have been available? One of the primary substantive rationales for stakeholder involvement is that the public brings a wealth of local knowledge about issues such as environmental conditions, land use, and exposure that can improve environmental assessments. In some cases, stakeholder groups even provided the impetus and resources for collecting new information. The Buffalo River Citizens Committee (BRCC), for example, was a major force behind better data collection on the water quality of the Buffalo River, which was part of an effort to clean up the area where the river joins Lake Erie. According to the researchers who analyzed BRCC's role in the cleanup, their work "led to a better environmental database on the river" whose information was

"co-produced and shared" between the BRCC and the New York Department of Environmental Conservation (Kellogg, 1993, 237). In 49 cases, participants added important information, and in 9 cases their failure to do so was noted.

Technical analysis. Did participants engage in technical analysis to improve the foundations on which decisions were based? Beyond providing information, participants can perform the analytical work of understanding problems, evaluating options, and identifying the likely results of different alternatives. For example, in the development of the Missouri River Basin Plan—concerning navigation and flood control on the river between Kansas City and St. Louis—stakeholders performed analyses on the economic, hydraulic, recreation, environmental, and land use aspects of various levee alternatives (Mazmanian and Nienaber, 1979, 75). In 44 cases, participants contributed to the technical analysis, and in 26 cases they did not.

Innovative ideas. Did stakeholders come up with innovative ideas? Stakeholder processes can be thought of as expanding the resources available for problem solving as many people approach the same problem from different perspectives. One example is a group of local ranchers and landowners in southern New Mexico and Arizona who came up with the idea of using protected grasslands as a "grass bank" to encourage conservation and discourage development of farmlands; under this plan, ranchers could use the grass on protected land in exchange for granting conservation easements on their own land (Bernard and Young, 1997). In 29 cases, participants were credited with contributing innovative ideas, and in 6 cases their failure to do so was noted.

Holistic approach. Did stakeholders develop a more holistic and integrated way of looking at an environmental problem? While agency personnel are often constrained by program mandates to look at problems in narrow ways, the public is not. The public's broader perspective can help define problems in ways that lead to more effective management. It can also broaden the opportunity for agreement among parties. Narrow water quality questions turn into watershed solutions; environmental cleanup decisions turn into economic development plans; resource permitting debates turn into comprehensive resource management planning. For example, in the case of a mediation regarding the damming of the Snoqualmie River in the 1970s, the question evolved from a yes/no question about building the dam to the question of "how do we provide some level of flood control, ensure the continued economic viability of the farmers and the towns, and build the kind of land use plans and controls that maintain the valley as a greenbelt with broad recreational value?" (Cormick and Patton, 1980, p. 88). Participants pushed decisionmakers to be more holistic in 25 cases, and their failure to do so was noted in 4 cases.

As a window on the outcomes of a varied set of stakeholder processes, the case study pool brought together here should lead us to an optimistic view of what such processes can accomplish. Across all of the direct and indirect criteria, considerably more cases appeared to produce good outcomes than bad. The aggregate criterion, which pulls together the varied definitions of quality and accounts for multiple criteria scores for a single case, reflects this balance in favor of good outcomes.

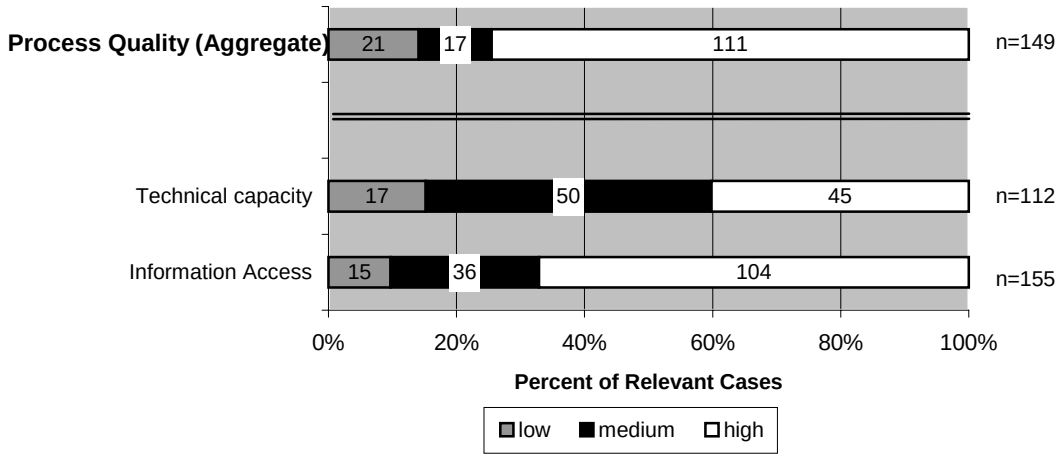
But outcomes are not the only way to judge the substantive quality of stakeholder decisions. We can also look to the process of participation and examine whether the scientific and technical resources available to stakeholders were adequate.

Substantive Quality of Process

To look at whether participatory processes provided adequate scientific and technical resources, two criteria were brought together into an aggregate measure. The first criterion evaluated the technical capacity of participants as defined by their training and experience with the issues under discussion. It can be thought of as a measure of "internal" technical resources. The second criterion evaluated participants' access to technical resources, in the form of information availability and access to expertise. It can be thought of as a measure of "external" technical resources.

The aggregate procedural quality measure and its two components are shown in Figure 5. A high score on the aggregate meant that the combination of internal and external resources provided participants with a relatively high level of technical resources. A low score meant that the process was deficient in internal and external technical resources. (The details on how the aggregate was constructed are included in Appendix A.) As shown in Figure 5, 149 cases out of 239 received an aggregate score. Nearly three times as many cases scored high as those that scored medium or low. The abundance of high-scoring cases suggests that the technical and scientific resources available to most of these cases were indeed quite good.

Figure 5: Process Quality



Note: Numbers within bars refer to the number of cases in that category.

The individual component criteria in the aggregate measure are described below:

Technical capacity. The technical capacity criterion looks at the scientific knowledge and technical training that stakeholders brought to the process. There is a tendency to assume that the citizens participating in environmental policy decisions are lay people rather than experts. Yet the capacity that participants bring to the table can often be quite impressive, both in terms of scientific and technical training and in terms of in-depth knowledge of the issues under discussion. An example of a highly skilled group is the Northern States Power Advisory Task Force, which included 2 physicists, a university biologist, other scientists and engineers, and many people with long histories of involvement in energy issues (Ducsik and Austin, 1986).

In roughly 40% of the cases for which data was available, there was a significant level of technical capacity among most of the participants. In another roughly 45%, there were at least some participants with significant technical capacity who could act as internal technical resources for the rest of the group. In the remaining cases, participants had little overt technical or issue-related expertise. It is only to this last 15% that the label "lay public" most appropriately applies.

Access to technical resources. The access to technical resources criterion looks at whether stakeholders had access to information and expertise that they felt were adequate and unbiased.

One of the long-standing concerns in public participation is that the public will have inadequate access to technical expertise or will have to rely too heavily on potentially biased information generated by agencies. Under some programs, agencies can provide technical assistance grants to public groups to help alleviate such imbalances of information. With or without agency funding, participants in the cases discussed here accessed expertise through a variety of methods, such as hiring consultants, interacting with technical advisory committees, or otherwise querying outside experts. An interesting model for these types of consultations were citizen juries, where a panel of citizens (the jury) listened to testimony and asked questions of a series of experts (the witnesses) in order to render informed judgement on a particular policy topic.

In 67% of the 155 cases for which data were available, participants had access to what they regarded as credible, relevant, and high quality technical information and expertise. In another 23%, there was some limited access to technical information and expertise.

Like the measures of outcome quality, the measures of process quality led to an optimistic view. On the individual criteria and the aggregate measures, processes with adequate technical and scientific resources far outweighed those that lacked them. The results seem to run counter to a concern raised by Yosie and Herbst (1998) that science and scientists were not well enough integrated into stakeholder processes.

Comparing Outcome and Process Measures

The two approaches to measuring substantive quality, one based on outcomes and one based on process, provide two perspectives on stakeholder-based decisionmaking. Surprisingly, the two measures were not highly correlated across cases.[3] Although there were many cases that scored high on both of the aggregate measures, unexpectedly, a number of cases scored high on one and low on the other. Why would quality outcomes not always be related to a quality process? Examining the cases where the two measures were not in agreement generates three insights into the criteria and the cases themselves.

The first answer is that the two measures are not reflecting the same dimensions of quality. The criteria that make up the process measure deal explicitly with scientific and technical dimensions of quality. The criteria that make up the outcome measure reflect a much

3 The correlation between the aggregate outcome quality measure and the aggregate process quality measure is 0.22 and is significant at the 90% confidence level (chi-squared probability=.10).

broader definition of quality. Some of the outcome criteria—such as cost-effectiveness, added information, or technical analysis—are arguably closely related to technical quality. But other criteria—such as joint gains or the opinion of participants—look at quality in terms that are much more subjective and dependent on the interests of the participants. If we are interested in scientific and technical quality specifically, we may want to look to the process quality measure. But if we are interested in a broader definition of quality, the outcome quality measure is more appropriate.

The second answer lies in the scope of the process being examined. The process measure is open-ended—it says nothing about actual decisions made. In a number of cases, highly trained participants or those with access to high quality technical resources did not have much of an impact on substantive outcomes because the design of the process gave them little ability to make or contribute to decisions.

The third answer deals with what the stakeholder-based decisions are being compared to. As outlined above, the outcome criteria incorporate a comparison to an implicit or explicit alternative or baseline. The direct quality measures are scored against a plausible alternative: for example, participants felt that decision X was more cost-effective or more satisfying to a broader range of interests than decision Y. Even the indirect quality measures assume an alternative: participants added information that otherwise would not have been available or lobbied for a more holistic approach that would not have been undertaken otherwise. The process criteria, on the other hand, are not scored relative to a baseline. In these cases, we often don't know what the alternative to a stakeholder process would have been.

The two substantive quality measures, then, give us two distinct ways of looking at substantive quality. One looks at the quality of outcomes, broadly defined. It takes into account the scope of a particular process and makes a comparison to alternatives. The second looks more explicitly at scientific and technical quality as process inputs but ignores the scope of the process itself, alternatives to the process, or the outcomes it might generate. Neither measure tells the whole story, but together they provide insights into the quality of stakeholder-based decisions. They also provide rankings that we can use to ask how different attributes of stakeholder processes affect quality.

5. The Importance of Process and the Lead Agency

The case studies examined here are quite heterogeneous, with wide differences across a number of characteristics, including the environmental issues under discussion, the types of participatory processes used, and the level of government at which decisions were made. Such differences among cases may affect the quality of the decisions stakeholders would make. This section examines two of these differences—the type of participatory process and the identity of the government agency leading it.

The methodology used here is a relatively simple process of stratifying the sample set based on a single attribute and comparing results between the two sets of cases. It uses a standard statistical test (a chi-squared test) to judge whether differences between the two sets are statistically significant or not. Using this approach has advantages in terms of ease of explication and is sufficient to illustrate the points made below. Its main weakness, however, is that it can mask the influence that unobserved factors are having on the results. The "unobserved factor" that is probably of principal concern here is the type of issue being addressed—whether it be land use planning, a hazardous waste cleanup, or the development of regulations. Preliminary data analysis using a more complex multivariate approach suggests that introducing considerations of issue type into the analysis does not have much impact on the results reported here. Instead of delving into more complex statistics, the discussion below addresses the multivariate nature of the data qualitatively, examining how results for the entire data set compare to those for subsets focused on specific environmental issues, such as hazardous waste cleanup and resource planning and management.

Type of Participatory Process

As outlined in Section 3, stakeholder processes can take on many forms, from a series of public meetings to an intensive negotiation among disputing parties. In Section 3, the cases were described in terms of four categories, each accounting for roughly a fourth of the data set. The first two categories did not involve consensus decisionmaking; these were public meetings and non-consensus advisory committees. The second two categories did involve consensus decisionmaking; these were consensus-based advisory committees and negotiations/mediations.

The consensus-based processes were generally more intense, and required more commitment from participants. Rather than simply expressing positions, participants were

seeking common ground. In the negotiation/mediation cases, participants were actually forging agreements. Recent years have seen a rise in these more intensive consensus-based forms of participation to inform and make environmental policy, including the use of federal advisory committees at contaminated DOE sites, national policy dialogues, and regulatory negotiations. EPA has embraced consensus-based stakeholder processes in many of its reinvention initiatives, such as Project XL and the Common Sense Initiative. At more local levels, consensus-based grassroots stakeholder councils have sprung up around the country to agree on how to manage natural resources (Weber, 2000).

It is the rise of the more intensive processes of consensus-building and agreement-seeking that have raised concern about the quality of decisions made, or influenced by, stakeholders. If the "political" element of stakeholder processes is indeed leading to a sacrifice in quality, then such a sacrifice should be most obvious in cases emphasizing consensus—and perhaps more so, those explicitly seeking an agreement among parties through negotiation and mediation.

Across both the outcome and process measures, however, the data suggest that negotiation and mediation, and consensus-seeking processes generally, score higher on quality criteria than less-intensive stakeholder processes. Figures 6 and 7 compare the four different types of participatory processes on the outcome quality and process quality measures. Looking first to outcome quality in Figure 6, the difference between negotiation/mediation cases (D) and the rest (A, B, and C) is positive and statistically significant.[4] The same is true if we compare the consensus-seeking group (C and D) with the non-consensus seeking group (A and B).[5] Looking next to process quality in Figure 7, the results are similarly positive and statistically significant but the magnitudes of difference are even higher.[6]

A subset of the data dealing with the cleanup of Superfund sites provides insight into the trends in the overall data. There are twenty-five Superfund cases in the dataset with information on the outcome or process measures. All of them involve EPA, and many involve DOE in the cleanup of its nuclear weapons facilities. The distribution of high, medium, and low scores for these Superfund cases across the outcome and process measures are roughly similar to the full

4 The chi-squared probability is .015.

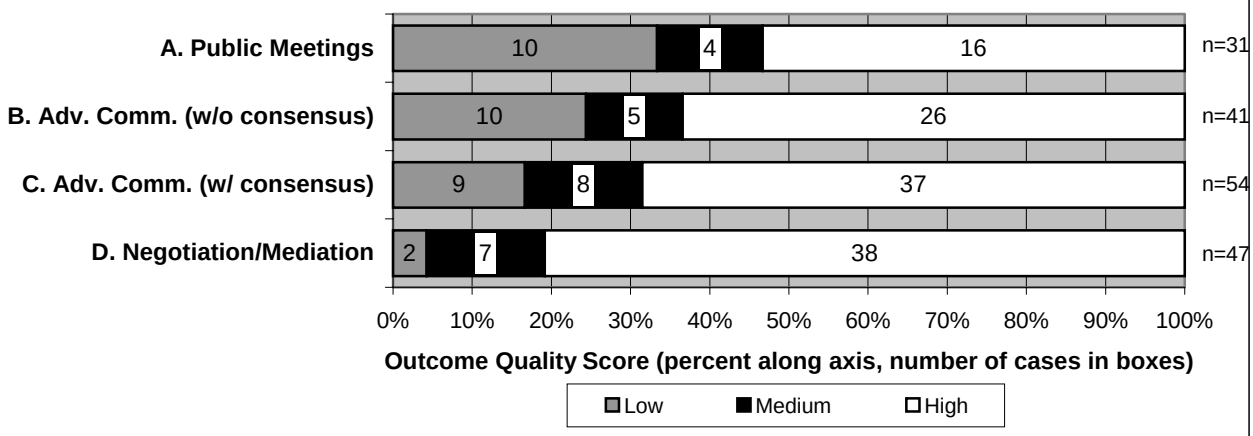
5 The chi-squared probability is .015.

6 In comparing type D to A, B, and C, the chi-squared probability is .009. In comparing types C and D to A and B, the chi-squared probability is .000.

data set. The relationship between substantive quality and process type also appears to hold up in the Superfund subset of the data. Brief descriptions of four cases serve to illustrate the results linking the type of process to substantive quality.

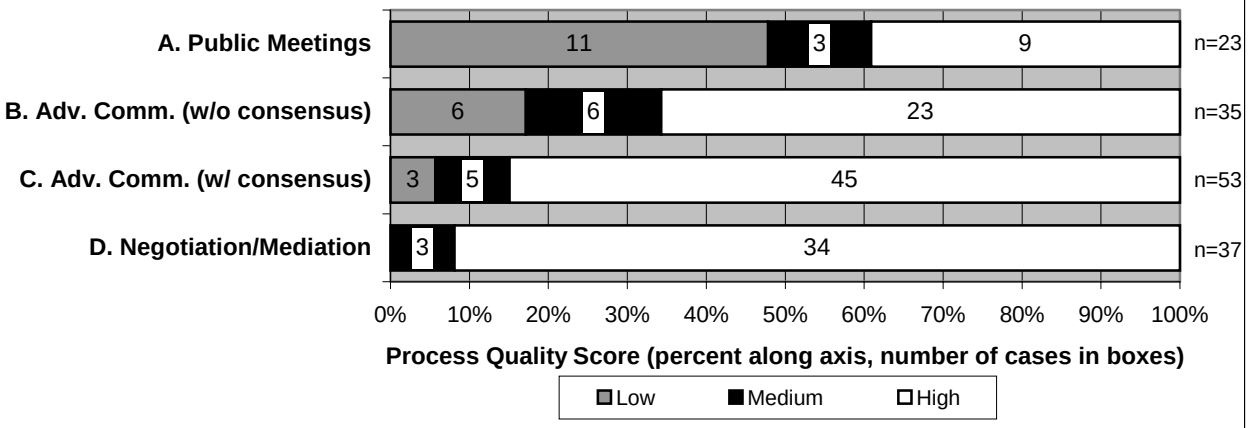
The two cases of the Lipari Landfill in Pittman, New Jersey, and Fort Ord near Monterey, California, illustrate how process can conspire to limit the public's contribution to the substance of decisions. At Lipari in the early 1980s, public meetings were the primary means to engage the public (Kauffman, 1992; Kaminstein, 1996). An agenda and scope tightly controlled by EPA, combined with few technical internal or external resources for participants, meant that local residents were effectively shut out of decisionmaking, and left feeling "ignorant and overwhelmed." (Kaminstein, 1996, 460). Accordingly, the local residents contributed little to the substance of decisionmaking. The situation in Fort Ord was similar. At Ford Ord, the participatory process was an advisory committee but one that explicitly did not seek consensus and was not intended to develop or recommend policies to the U.S. Army, who was cleaning up the site (Wernstedt and Hersh, 1997). Like the Lipari cases, participants had little technical training and little access to other resources, which effectively kept them out of any technical decisionmaking. In both the Lipari and Fort Ord cases, the participatory process was simply not robust enough to allow participants to develop ideas, share information, and formulate alternatives that might promise to improve decisionmaking. In both cases, cleanup decisions remained mired in controversy.

Figure 6: Outcome Quality and Mechanism Type



Note: Numbers within bars refer to the number of cases in that category.

Figure 7: Process Quality and Mechanism Type



Note: Numbers within bars refer to the number of cases in that category.

The Lipari and Fort Ord cases stand in stark contrast to two other cases of decisionmaking about hazardous waste sites—the cleanup of a hazardous waste site in New Bedford, Massachusetts, in the early 1980s and the remediation of a contaminated DOE facility in Fernald, Ohio, in the early 1990s. At New Bedford, years of controversy over whether or not to incinerate contaminated material were settled by a consensus-based stakeholder process called the New Bedford Harbor Forum (Hartley, 1998; Hartley, 1999). The forum brought together a group of local residents and officials, all with varying degrees of expertise in the issues. Citizens hired a technical advisor and got involved in the technical analysis; they contributed local information about the site. The deliberations generated a more holistic cleanup solution that incorporated the role of the site in the economic revitalization of the community. The process ultimately resulted in what the case study author called "better decisions that reduced risk" (Hartley, 1998, p. 6). The experience in Fernald, Ohio, was similarly successful (Applegate, 1998; Duffield and Depoe, 1997). There, DOE established a consensus-based advisory committee to make decisions about complex and intertwined issues, such as on-site or off-site disposal, future use of the site, the acceptable level of residual risk, and appropriate cleanup technologies. In two years of work, utilizing both the internal expertise of some of the members of the committee and external consultants, the participants arrived at what was considered to be a fair and balanced cleanup strategy, which DOE regarded as faster, cheaper, and more holistic than what DOE would have developed alone. Noted earlier in this paper is the fact that DOE considers the outcome to have saved taxpayers $2 billion.

The pursuit of consensus through deliberation is the defining feature of these successful hazardous waste cases, and the higher-scoring cases in the larger data set more generally. Of course these kinds of cases are often longer, better funded, and attract more committed participants than less intensive decision processes. All of these factors are all likely reflected in the result that links consensus-based processes with higher quality. But there are reasons to think that consensus-seeking plays a more direct role in supporting decision quality. Resolving conflict often requires dealing with scientific uncertainty through appeals to independent expertise, joint-fact finding on the part of all participants, or new research altogether. Arguments are generally won or lost based on the quality of the information. Mistrust among stakeholders and between stakeholders and government may uncover questionable science and bad ideas. Building trust may require tapping into independent sources of expertise or generating new knowledge. All of these suggest that the "political" features of these more intensive stakeholder processes may create a positive synergy with the quality of its outcomes.

The Nature of the Lead Agency

Lead agencies play a large in role in the design and execution of stakeholder processes. So too might the locus of decisionmaking, whether it is national, state, or local in scope. We examine two questions related to the nature of the lead agency. First is whether processes led by state and local governments compare favorably to processes led by federal agencies. Second is how processes led by the EPA compare to both federal agencies alone and to agencies at all levels of government. The results of the analysis are shown in Figures 8 and 9 and are discussed below.

One of the surprising aspects of the research described here is the large number of cases of stakeholder processes undertaken at the state and local level. State and local decisionmaking is likely to increase in importance as local issues, such as land use, come to the forefront of environmental concern. Stakeholder participation at the state and local level is likely to increase as well. In fact, the National Governor's Association has adopted collaboration among stakeholders as one of the core principles of its "Enlibra" doctrine, which outlines a vision for environmental policymaking in the states. A shift toward participatory decisionmaking at the state and local levels could be cause for concern from the point of view of quality, particularly if fewer scientific and technical resources are available at these levels.

As indicated by Figures 8 and 9, however, there is not much of a difference between the results for state and local lead agencies as compared to federal agencies for either quality measure.[7] For example, on the aggregate outcome measure there was a higher percentage of high scoring cases led by state and local governments, but there was also a higher percentage of low scoring cases. To see how the results play out for specific types of issues, we briefly examine the set of cases dealing with resource planning and management.

State and local governments play a large role in developing policies regarding how resources will be used and managed. There are 74 cases in the data set dealing with resource planning and management, encompassing land use decisions, habitat conservation planning, watershed management, the management of parks and other designated areas, water quality planning, and wildlife management. Nearly all of the cases dealt with specific sites or geographic regions rather than overarching policy issues. Of these 74 cases, 38 were led by state

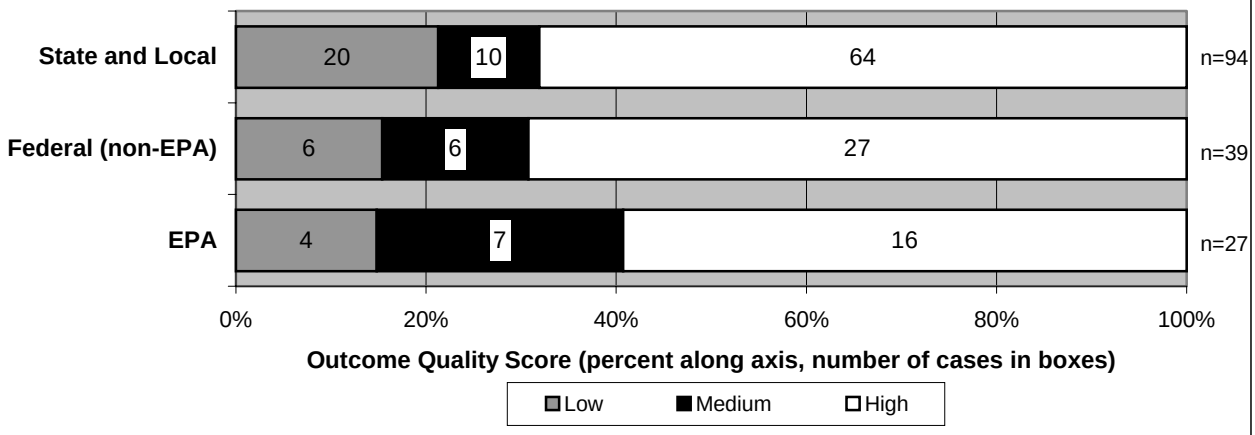
7 On neither measure is the difference between state/local agencies and federal agencies statistically significant. On the outcome measure, the chi-squared probability is .219. On the process measure it is .972.

and local agencies while 26 were led by federal agencies. The other ten were jointly led by state and federal agencies or some other arrangement. Like the data set as a whole, state and local agencies performed about the same as federal agencies in resource planning and management. High scoring cases made up 60% to 80 % of the sample, and low scoring cases made up to 10% to 30%, regardless of the level of the lead agency. The fact that state and local agencies do not appear to be doing worse on quality measures—either across the data set or for resources planning and management specifically—should help quell concerns that increased state and local stakeholder decisionmaking may sacrifice quality.

A second issue related to the nature of the lead agency is whether processes led by the EPA compare favorably to those led by other federal, state, and local agencies. Because some of the recent concern over the role of such processes has centered on EPA, it makes sense to single the agency out for analysis. EPA has embraced stakeholder-based processes in many of its reinvention initiatives but has also been the target of criticism for how it handles participation in some other programs, such as Superfund (GAO, 1994).

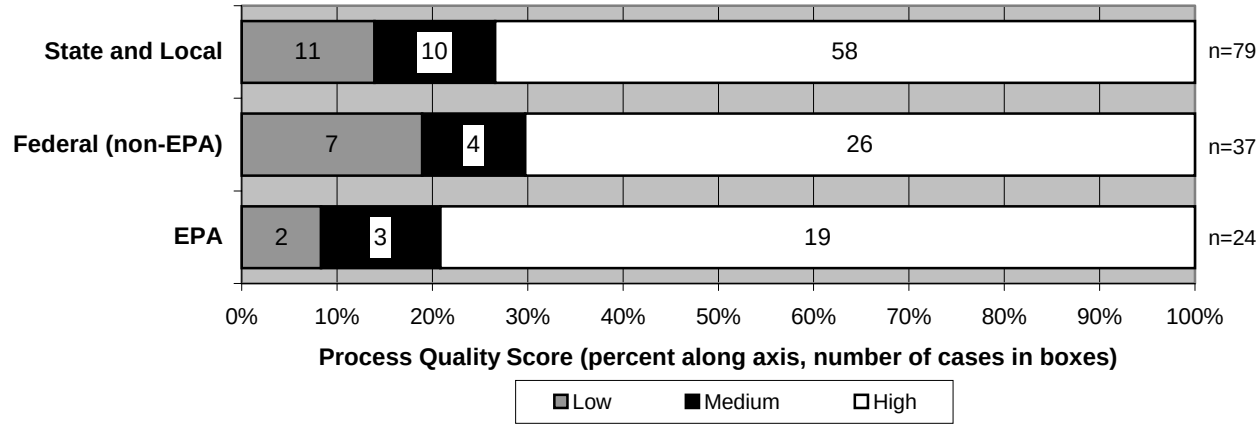
Thirty-two of the 239 cases involved EPA as the primary lead agency; it was represented the most out of all single agencies in the set of cases. The EPA cases in the dataset tended to be relatively high profile. Fourteen dealt with oft-contentious Superfund cleanups and related issues, seven concerned regulatory negotiations, and another three dealt with the EPA reinvention program Project XL. The remaining eight covered a variety of other regulatory and policymaking arenas.

Figure 8: Outcome Quality and Government Type



Note: Numbers within bars refer to the number of cases in that category.

Figure 9: Process Quality and Government Type



Note: Numbers within bars refer to the number of cases in that category.

As with state and local governments, there was little difference between EPA-led cases and those led by other federal agencies or agencies at all levels of government.[8] Even taking into account the relatively high-profile kinds of cases that EPA gets involved in, the agency's stakeholder processes appear to be doing no better or worse than the norm.

The analysis of the relationship between the nature of the lead agency and the quality of stakeholder-based decisions suggests that it is not the place to look for explaining differences in the quality of stakeholder-based decisions. Processes led by state and local governments do not appear to be more or less successful in quality terms than their federal counterparts, even for the subset of the data focusing on resource planning and management. Of particular interest, they don't seem to result in less technical capacity or less access to scientific and technical resources. Likewise, processes led by EPA do not appear to result in higher or lower quality decisions than processes led by other agencies. All agencies likely face similar challenges in developing processes that adequately incorporate scientific and technical resources and that support quality outcomes.

6. Conclusion

Based on an examination of 239 case studies, we should be rather optimistic about the quality of stakeholder-based decisions. Even though the data are not as systematic and complete as the ideal, and the case study record may be an imperfect window on the world of stakeholder processes, the analysis should give some reassurance that the "political" aspects of stakeholder processes are not sacrificing decision quality. Across a diversity of process types, levels of government, and environmental issues, most of the evidence points toward quality decisionmaking from stakeholder processes.

The analysis should help clarify, as well, how to think about the quality of stakeholder-based decisions. The outcome criteria sketch a broad and complex understanding of quality. The process criteria provide more narrow insights into scientific and technical quality. In fact, it may be the only perspective into how stakeholder processes utilize science, given the difficulty

8 None of the differences were statistically significant. Compared with only federal agencies, the chi-squared probability for the outcome measure was .564 (n=66) and for the process measure was .523 (n=61). Compared with all local, state, and federal agencies, the chi-squared probability for the outcome measure was .168 (n=160) and for the process measure was .66 (n=140).

of identifying an outcome measure that can somehow measure the "scientific quality" of a decision.

Beyond the direct question of the quality of stakeholder-based decisions, there are two other issues that ought to be considered in thinking about the impacts of stakeholder processes on environmental policymaking. They deserve at least passing mention here.

The first question is whether stakeholder decisions are being implemented. There should be far less concern about the quality of stakeholder decisions if administrative and political checks and balances are in place to halt bad decisions on the road to implementation. At a very basic level, agencies rather than stakeholders usually implement decisions, creating a strong filter between stakeholder decisions and action. Indeed, various studies of implementation suggest that agreements among participants do not necessarily translate directly into actual impact on policy. In a study of mediated environmental issues, Bingham (1986) noted an important gap between agreements among mediating parties and the implementation of those agreements. In research on regulatory negotiations, Coglianese (1997) found that agreements reached through negotiations were often revisited after subsequent controversy. In an earlier study by the author, no consistent link between good public participation and implementation could be identified in a series of cases of stakeholder-based planning in the Great Lakes region (Beierle and Konisky, 2000). Another analyst of the same Great Lakes cases concluded that whatever implementation had occurred had very little to do with the stakeholder planning process (Gurtner-Zimmermann, 1996). While much more work on the relationship between participation and implementation needs to be done, there is much evidence to suggest that various checks and balances on stakeholder-based decisions are solidly in place—for better or worse.

The second question is what we should be comparing stakeholder processes to. There is a tendency to contrast stakeholder processes with more expert-led scientific decisionmaking—another chapter in the long running debate about whether pluralism or managerialism should inform agency discretion (Stewart, 1975; Reich, 1985). Yet studies of agency decisionmaking suggest that the status quo to which stakeholder processes are an alternative are not these more idealized technocratic approaches. Rather, agency decisionmaking is already quite "political," subject to partisan winds and interest group influence. Charnley (2000) points out that it is just such criticisms of the status quo that have raised concerns about the use of science in environmental decisionmaking in the first place. Dissatisfaction with the status quo is one of the primary reasons that stakeholder processes are on the rise. Finally, many decisions made by agencies, even EPA, are not really about science at all (Powell, 1999). Stakeholder processes may work to improve on traditional agency decisionmaking by making processes more

formalized and transparent and by giving non-scientific issues the “political” hearing they should appropriately have.

Evidence about the quality of stakeholder decisions, the presence of checks and balances in the implementation process, and the less-than-stellar status quo come together as a strong endorsement for stakeholder-based decisionmaking.. There may be many ways to produce decisions of high technical quality, but there are relatively few that do so while also educating the public, eliciting public values, resolving conflict, and building trust in agencies, as many stakeholder processes do. That we can make some headway on these more “political” features of decisionmaking while not also sacrificing quality is indeed a positive endorsement for opening the doors of decisionmaking to the public.

Appendix A: Construction of Aggregate Measures

Outcome Quality Aggregate

The outcome quality aggregate combines data from the eight component quality criteria listed in Figure 4 of the paper. The range of eight quality criteria reflected the different aspects of "quality" relevant to different kinds of cases. For example, one could judge a mediation case on the basis of whether the decisions reached increased joint gains for those involved. A public meeting, however, would be more appropriately judged on whether participants contributed information or ideas that would not otherwise have been available.

Because different criteria were appropriate for different kinds of cases, it was quite rare to have more than two criteria scored for each case. In fact, out of 172 cases—for which at least one of the eight criteria were scored—none were scored on more than five criteria, and only 47 were scored on three to five criteria. Fifty-three were scored based on two criteria; and 72 were scored based on only one criterion. The lack of substantial overlap among the quality criteria meant that it was problematic to inter-correlate them in order to see whether they described some overarching conceptual meaning of "quality." As shown in Table A1, the largest number of pairwise comparisons that could be made was for 32 cases; and for many pairs of criteria, correlation coefficients could not be calculated because there was no variation in one of the criterion. Nevertheless, some of the eight criteria appear to hang together quite well.

Table A1 Inter-correlation of Outcome Quality Criteria

	cost-effective	joint gains	opinion	other direct	infor-mation	technical analysis	innova-tive ideas	holistic perspec-tive
cost-effective	1							
joint gains	-.17 (n=11)	1						
opinion	.49 (n=7)	.64 (n=18)	1					
other direct	1.0 (n=4)	-.37 (n=11)	*6/8 agree	1				
informa-tion	.29 (n=4)	.60 (n=13)	*6/8 agree	*4/4 agree	1			
technical analysis	.29 (n=4)	*12/13 agree	.82 (n=10)	*4/4 agree	.84 (n=32)	1		
innova-tive ideas	.63 (n=6)	*14/18 agree	.65 (n=8)	*2/2 agree	.85 (n=13)	1.0 (n=16)	1	
holistic perspec-tive	.13 (n=6)	*7/9 agree	*9/9 agree	*1/2 agree	1.0 (n=12)	1.0 (n=11)	1.0 (n=10)	1

*No coefficient calculated because of no variation in one of the criteria. The ratio of agreements to total is reported.

If not describing a stand alone concept of "quality" derived statistically, the aggregate measure is at least an accurate reflection of the underlying quality criteria. The rules used to construct the aggregate were as follows:

1. For indirect quality measures, "yes" and "no" scores were converted into "high" and "low" scores, respectively.
2. If there were no low/high combinations for a given case (suggesting a wide divergence of scores), the scores were averaged (using 3, 2, and 1 for high, low, and medium, respectively). For cases with low/high combinations, we skipped to step 4.

3. Averages were rounded to the nearest score. Where the average fell exactly between two scores (e.g., 2.5) it was rounded up to the higher score.
4. Where there were low/high combinations, scores were determined on a case by case basis.

In developing an aggregate measure for 172 cases, 156 (90%) of them could be scored after step 2. This means that where there were multiple scores, they were all in agreement. Seven cases required averaging (step 3), but only one of these had to be rounded up from an average score that fell midway between two scores. For 9 cases, the aggregate was determined on a case-by-case basis (step 4)—all of these cases were given a "medium" based on mixed results.

Process Quality Aggregate

The process quality aggregate was constructed from two measures: the technical capacity of participants and their access to technical resources. In developing the aggregate, the following rules were used:

1. Cases scored high if at least one of the criteria were scored high and the other medium. This meant that high quality internal or external resources were present without an off-setting lack of one or the other.
2. Cases scored medium if both criteria were scored medium or if there were high/low combinations. This meant that either internal and external resources were moderate or that a high level of one of the two was offset by a low level of the other.
3. Cases scored low if at least one of the criteria was scored low and the other medium. This meant that the process was deficient in internal or external resources and the deficiency was not compensated by either internal capacity or external access to information
4. Data in which only one of the two components were scored were given a score based on the most likely score for the missing criterion.

The data fell into categories outlined in Table A2.

Table A2 Combinations of Process Quality Criteria Used to Construct Aggregate

Technical Capacity **Access to Technical Resources**	**no score**	**low**	**medium**	**high**
no score	No score (n=65)	Assumed **Low** (n=2)	No score could be assumed (n=9)	Assumed **High** (n=8)
low	Assumed **Low** (n=6)	**Low** (n=6)	**Low** (n=3)	**Med.** (n=0)
medium	No score could be assumed (n=16)	**Low** (n=4)	**Med.** (n=12)	**High** (n=4)
high	Assumed **High** (n=40)	**Med.** (n=5)	**High** (n=26)	**High** (n=33)

Appendix B: Examination of Bias

As mentioned in Section 2 of the paper, the potential for bias in a case survey is always of concern. This appendix deals with three possible sources of bias, all of which would make the pool of cases look more successful than the norm. The first potential source of bias arises if case study authors are more likely to write up successful cases than unsuccessful ones. The second potential source of bias arises if certain kinds of case study authors have a particular interest in making a case appear more successful than it really was. The third potential source of bias relates to the coding of the indirect quality measures. The first two can be treated together and the third treated separately.

Potential Bias in Case Selection and Description

While coding the cases, researchers flagged cases that 1) were picked to explicitly illustrate a successful or unsuccessful process or 2) were written by someone closely affiliated with the case (such as a participant or lead agency staffer) who might have an incentive to over-emphasize the case's success. Out of a total of 239 cases, 70 were picked to illustrate successful or unsuccessful cases, 150 were not, and in 19 it was not known. Sixty-six cases were written by someone closely affiliated with the case, 114 were not, and it was unknown for 59. Combining the two sets of cases, where the presence of one or both possible sources of bias made the case potentially biased, we were left with 118 potentially biased cases, 87 unbiased cases, and 34 in which it was unknown. Figure B1 compares the potentially biased cases with the unbiased cases across both the outcome quality and process quality measures. Although the potentially biased cases appear to be slightly more successful, the difference is small and not statistically significant for either the outcome measures or the process measures.[9] These two sources of potential bias, then, do not appear to be having much of an impact on the results described in this paper.

Potential Bias in Indirect Quality Measures

The third possible source of bias deals with the indirect quality measures. The indirect quality criteria describe what the participants did or did not do. Did they add new information or did they not? Did they come up with innovative ideas or did they not? It may be that case study

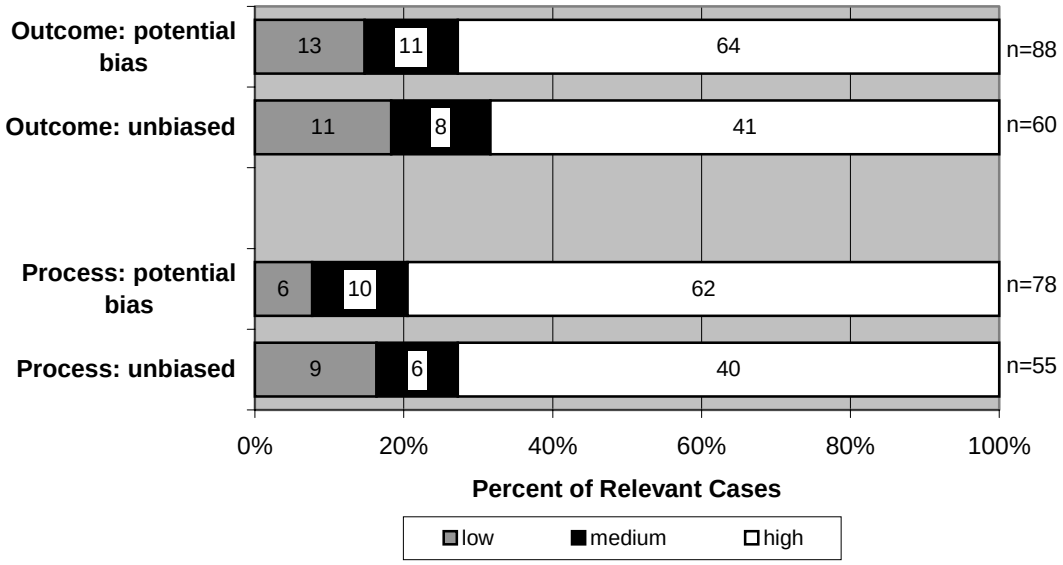
9 For the outcome measure, the chi-squared probability is .821 (n=148). For the process measure, the chi-squared probability is .295 (n=133).

authors were more likely to report positive information on these criteria than negative information. In reading through a case, coders could tell when stakeholders were adding information or doing analyses, for example, but were never quite sure what was going on if case study authors didn't report on these kinds of activities. Where they not done or were they just not deemed important by the case study author? In short, we can never be sure how to interpret gaps in the data.

Without more information on the actual cases, it is difficult to definitively tell whether the indirect criteria results are biased toward positive information or not. What we can do is compare the results for the indirect criteria with the results for the direct criteria, which are not subject to the same potential bias. Out of 172 cases with an outcome measure score, 69 were based only on indirect criteria and 103 could be recalculated to be based only on direct criteria. Figure B2 compares the direct and indirect criteria scores. Rather than being more positive, as would be the case if the indirect scores were biased, the indirect scores actually appear to be more negative (i.e., a higher percentage of low scores), and the difference between the two sets of cases is statistically significant.[10] The indirect criteria do not appear, then, to be adding a success bias to the results described in this paper.

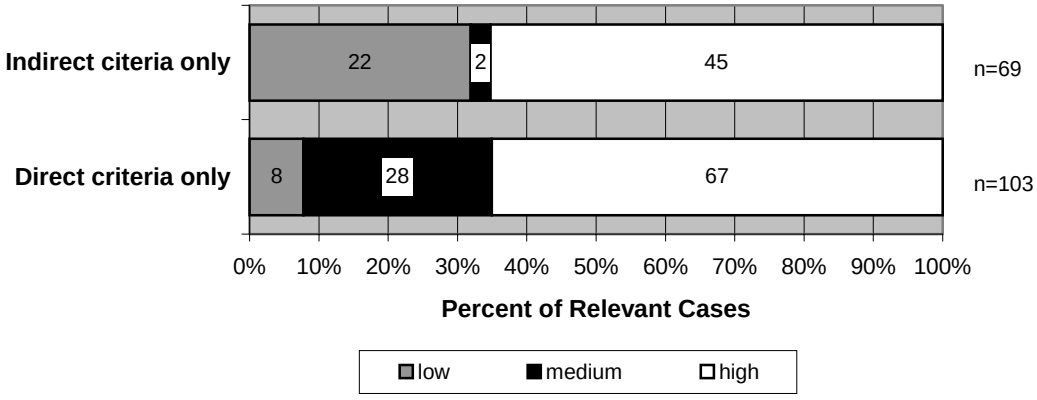
10 The chi-squared probability is .000. The significance of the difference between the two sets of cases can probably be explained by the low number of "medium" scores for the "indirect criteria only" category. This is not surprising because indirect criteria were only coded as "no" (translated as "low") and "yes" (translated as "high"). The two medium cases reported in Figure B2 come from the few cases where one indirect criteria was coded as "no" and another was coded as "yes."

Figure B1: Examination of Selection and Author Bias



Note: Numbers within bars refer to the number of cases in that category.

Figure B2: Examination of Indirect Criteria Bias



Note: Numbers within bars refer to the number of cases in that category.

References

Applegate, J. 1998. Beyond the usual suspects: the use of citizens advisory boards in environmental decisionmaking, *Indiana Law Journal*, 73(3), 1-43.

Beierle, T. & Konisky, D. 2000. Values, conflict, and trust in participatory environmental planning, *Journal of Policy Analysis and Management*, 19(4), 587-602.

Bernard, T. & Young, J. 1997. The ecology of hope: Communities collaborate for sustainability. East Haven, CT: New Society Publishers.

Bingham, G. 1986. Resolving environmental disputes: A decade of experience. Washington, DC: The Conservation Foundation.

Bullock, R. & Tubbs, M. 1987. The case meta-analysis method for OD. *Research in Organizational Change and Development*, 1, 171-228.

Charnley, G. 2000. Democratic science: Enhancing the role of science in stakeholder-based risk management decision-making. Washington, DC: Health Risk Strategies.

Coglianese, C. 1997. Assessing consensus: The promise and performance of negotiated rulemaking. *Duke Law Journal*, 46(6), 1255-1349.

Cormick, G. & Patton, L. 1980. Environmental mediation: Defining the process through experience. In Laura Lake ed., *Environmental Mediation*. Boulder, CO: Westview Press.

Daisey, J. & Morgan, G. 1999. Science Advisory Board commentary on the role of science in "new approaches" to environmental decisionmaking that focus on stakeholder involvement. Letter to Carol Browner, Administrator, Environmental Protection Agency, EPA-SAB-EC-COM-00-002, (October 7).

Ducsik, D. & Austin, T. 1986. Open power plant siting: The pioneering (and successful) experience of Northern States Power public involvement. In Dennis Ducsik, ed., *Energy Facility Planning: The Electric Utility Experience*. Boulder, CO: Westview Press.

Duffield, Jennifer & Stephen Depoe. 1997. Lessons from Fernald: Reversing NIMBYism through Democratic Decision-Making, *Risk Policy Report* (February 21).

Federal Facilities Environmental Restoration Dialogue Committee (FFERDC). 1996. Consensus Principles and Recommendations for Improving Federal Facilities Cleanup Washington, DC: Environmental Protection Agency (April).

General Accounting Office (GAO). 1991. Designing evaluations. Washington, DC: General Accounting Office.

General Accounting Office. 1994. Superfund: EPA's Community Relations Efforts Could Be More Effective, Washington, DC: General Accounting Office GAO/RCED-94-156.

Gregory, R. 2000. Using stakeholder values to make smarter environmental decisions. *Environment* 24(5), 34-44.

Gurtner-Zimmermann, A. 1996. Analysis of Lower Green Bay and Fox River, Collingwood Harbour, Spanish Harbour, and the Metro Toronto and Region Remedial Action Plan (RAP) process. *Environmental Management*, 20(4), 449-459.

Hartley, Troy. 1998. Participant Competencies in Deliberative Discourse: Cases of Collaborative Decision-Making in the U.S. EPA Superfund Program, presented at Seventh International Symposium on Society and Natural Resource Management, University of Missouri-Columbia (May 27-31).

Hartley, Troy. 1999. Versatility, Patience & Non-Defensiveness: Helping Participants Cope in Collaborative Environmental Decision-Making, draft manuscript submitted to *Negotiation Journal* (March).

Kaminstein, Dana S. 1996. Persuasion in a Toxic Community: Rhetorical Aspects of Public Meetings. *Human Organization*, 55(4), 458-464.

Kauffman, Stephen E. 1992. The Causes of Conflict and Methods of Resolution in a Citizen Participation Program: A Case Study of the Lipari Landfill Superfund Site, Ph.D. Dissertation, Bryn Mawr College.

Kellogg, W. 1993. Ecology and community in the Great Lakes basin: the role of stakeholders and advisory committees in the environmental planning process. Doctoral dissertation, Cornell University.

Larrson, R. 1993. Case survey methodology: Quantitative analysis of patterns across case studies. *Academy of Management Journal*, 36(6), 1515-1546.

Lucas, W. 1974. The case survey method: aggregating case experience (Rand Report No. R-1515-RC). Washington, DC: Rand Corporation.

Mazmanian, D. & Nienaber, J. 1979. Can Organizations Change? Environmental Protection, Citizen Participation, and the Army Corps of Engineers. Washington, DC: Brookings Institution.

Perhac, R. 1996. Defining risk: Normative considerations. *Human and Ecological Risk Assessment* 2(2), 381-392.

Perhac, R. 1998. Comparative Risk Assessment: Where Does the Public Fit In? *Science, Technology & Human Values*. 23(2), 221-242.

Powell, M. 1999. *Science at EPA: Information in the Regulatory Process*. Washington, DC: Resources for the Future.

Reich, R. 1985. Public administration and public deliberation: an interpretive essay. *Yale Law Journal*, 94(7), 1617-1641.

STATA. 1997. STATA Reference Manual, Release 5, Volume 3, P-Z. Stata Press: College Station, Texas.

Stewart. R. 1975. The reformation of American administrative law, *Harvard Law Review* 88, 1669-1813.

Stokes, Maura E., Charles S. Davis, & Gary G. Koch. 1995. *Categorical Data Analysis Using the SAS System*. North Carolina: SAS Institute, Inc.

Weber, E.P. 2000. A new vanguard for the environment: Grass-roots ecosystem management as a new environmental movement. *Society & Natural Resources* 13, Part 3, 237-260.

Wernstedt, Kris and Robert Hersh. 1997. Land Use and Remedy Selection: Experience from the Field—The Fort Ord Site, RFF Discussion Paper 97-28, Washington, DC: Resources for the Future.

Western Center for Environmental Decision-Making (WCED) 1997. Public involvement in comparative risk projects. Boulder, CO: Western Center for Environmental Decision-Making.

Yin, R. & Heald, K. 1975. Using the case survey method to analyze policy studies. *Administrative Science Quarterly*, 20, 371-381.

Yosie, T. & Herbst, T. 1998. Using stakeholder processes in environmental decisionmaking: an evaluation of lessons learned, key issues, and future challenges. Washington, DC: Ruder Finn Washington.